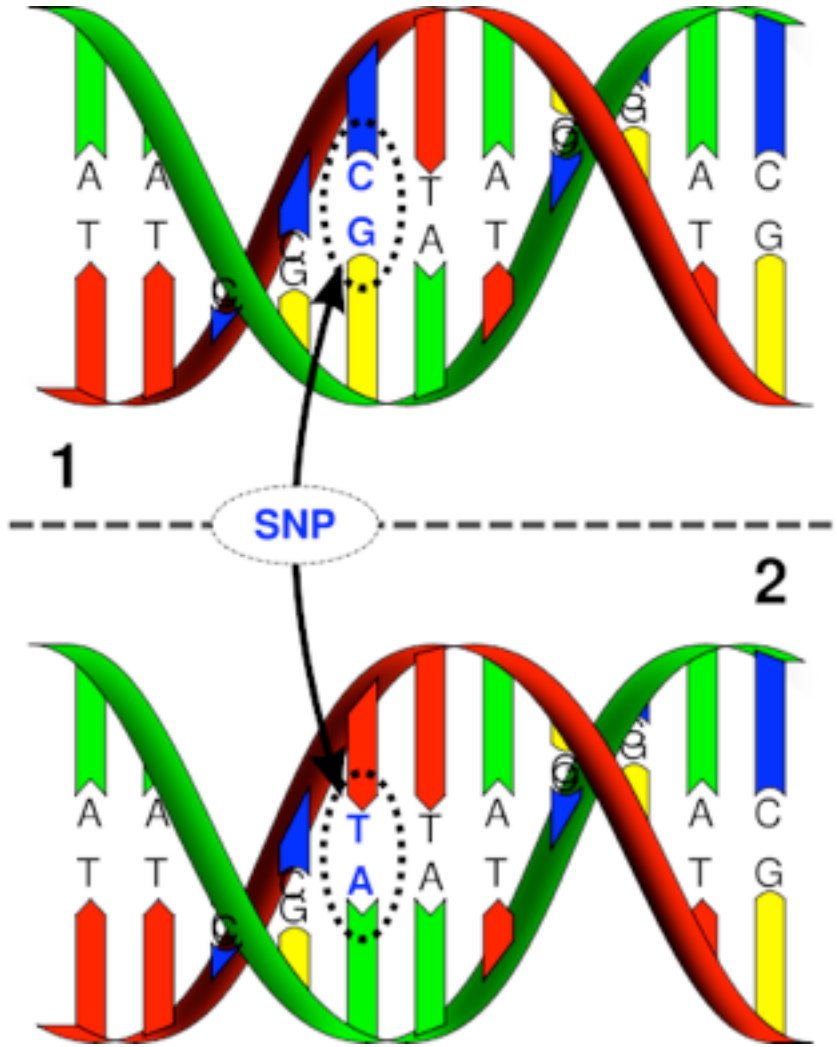
Single Nucleotide Polymorphisms (SNPs)



CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Sequence variations

- Single nucleotide polymorphisms
- Insertions/deletions
- Copy number variations (large: >1kb)
- Variable (short) number tandem repeats

DTU

CENTERFOR BIOLOGICAL SEQUENCE ANALYSIS CBS

Single Nucleotide Polymorphisms (SNPs)

- A single nucleotide (A,T,C,G) DNA sequence alteration

  ... **ACGGCTAA** ...

  ... **ATGGCTAA** ...

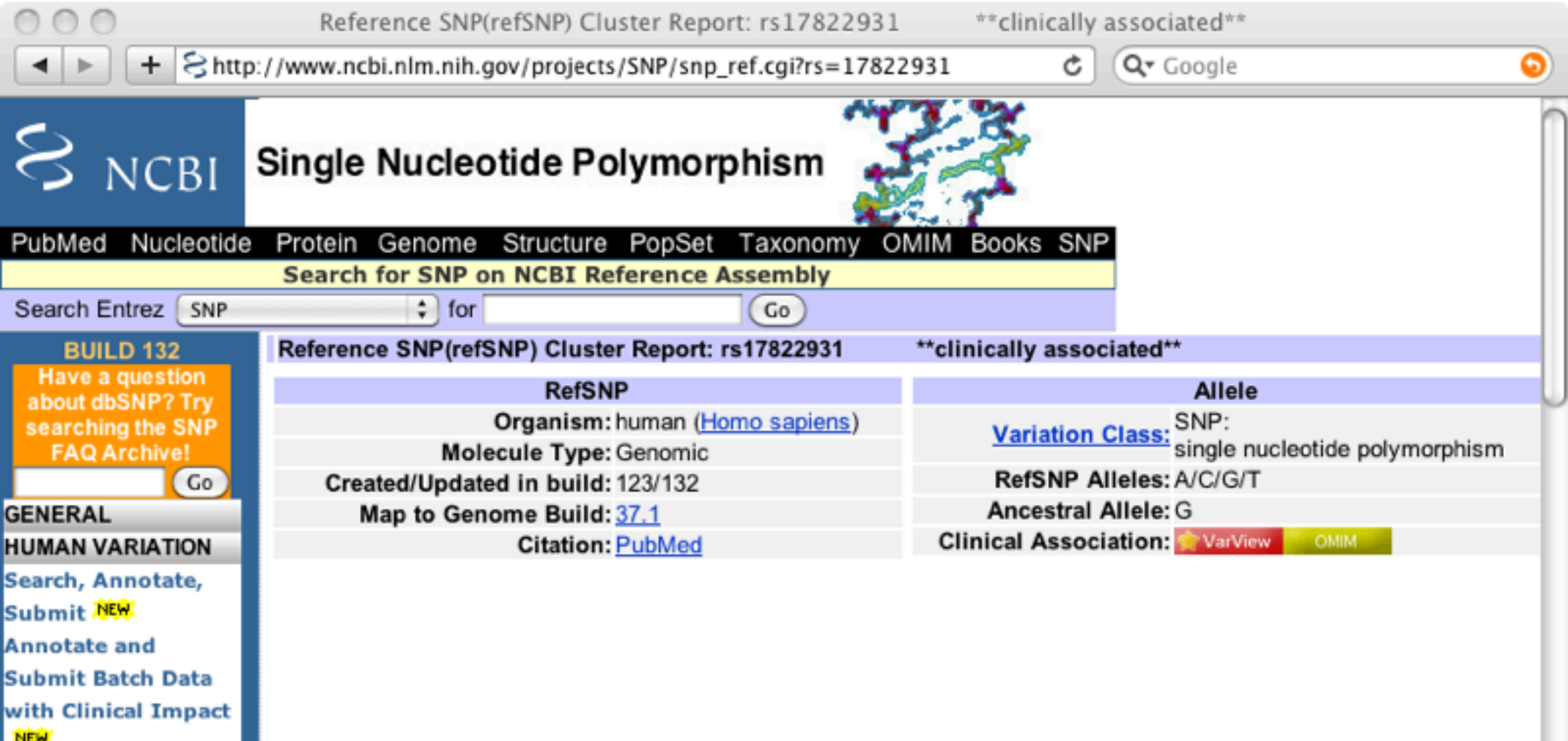
- It must occur in at least 1% of the population
- ~30 million SNPs
- SNPs make up ~80-90% of all human genetic variations
- Occur every 100-300 bases along the 3-billion-base human genome
- Evolutionary stable

DTU

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

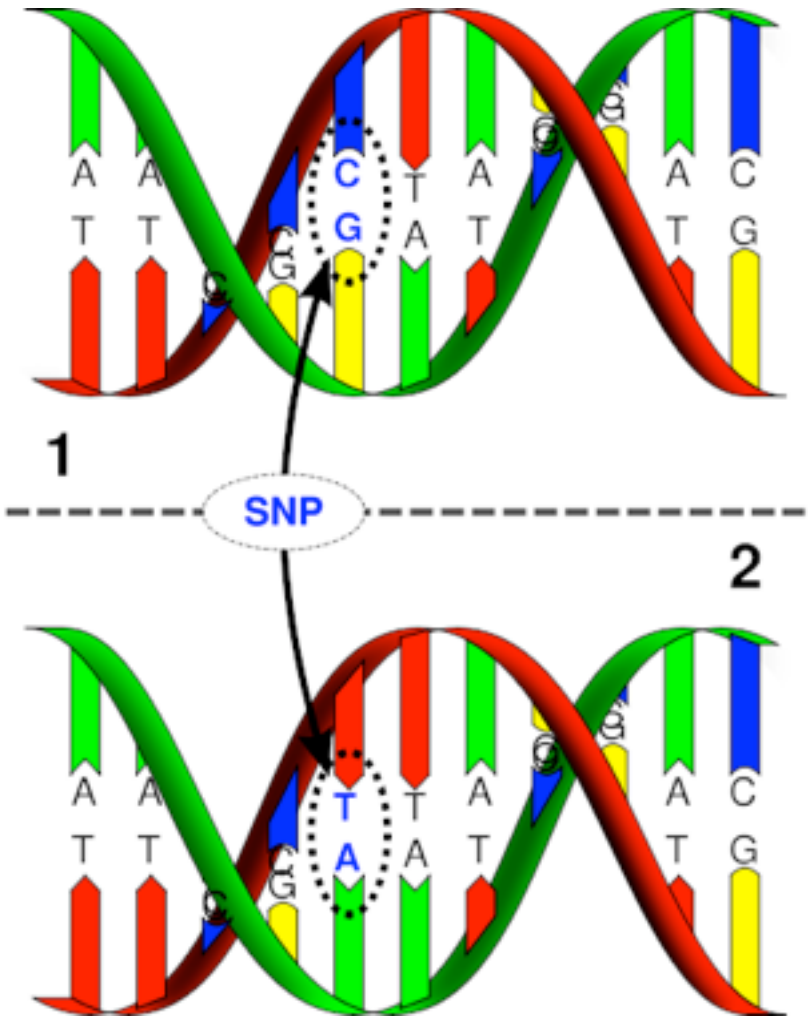
dbSNP

- rs numbers
- chromosome and positions (ncbi36 vs GRCh37)



Strand issues

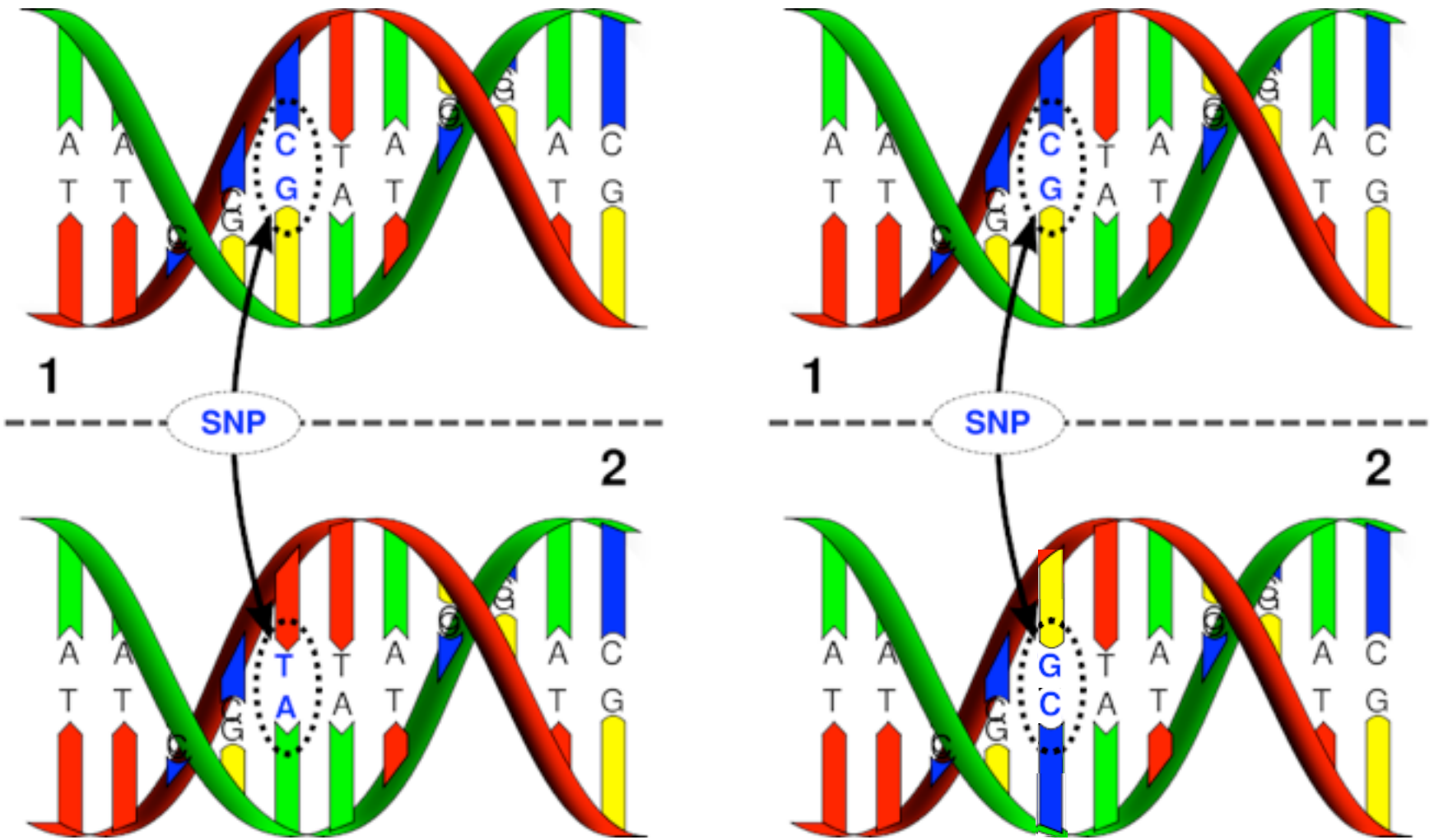
Importance of noting whether a SNP is mapped to forward or reverse strand



CENTERFOR BIOLOGICAL SEQUENCE ANALYSIS CBS

Strand issues

Importance of noting whether a SNP is mapped to forward or reverse strand



DTU

SNP arrays

Affymetrix SNP 6.0

Affymetrix SNP 5.0

Affymetrix GeneChip Human Mapping 250K

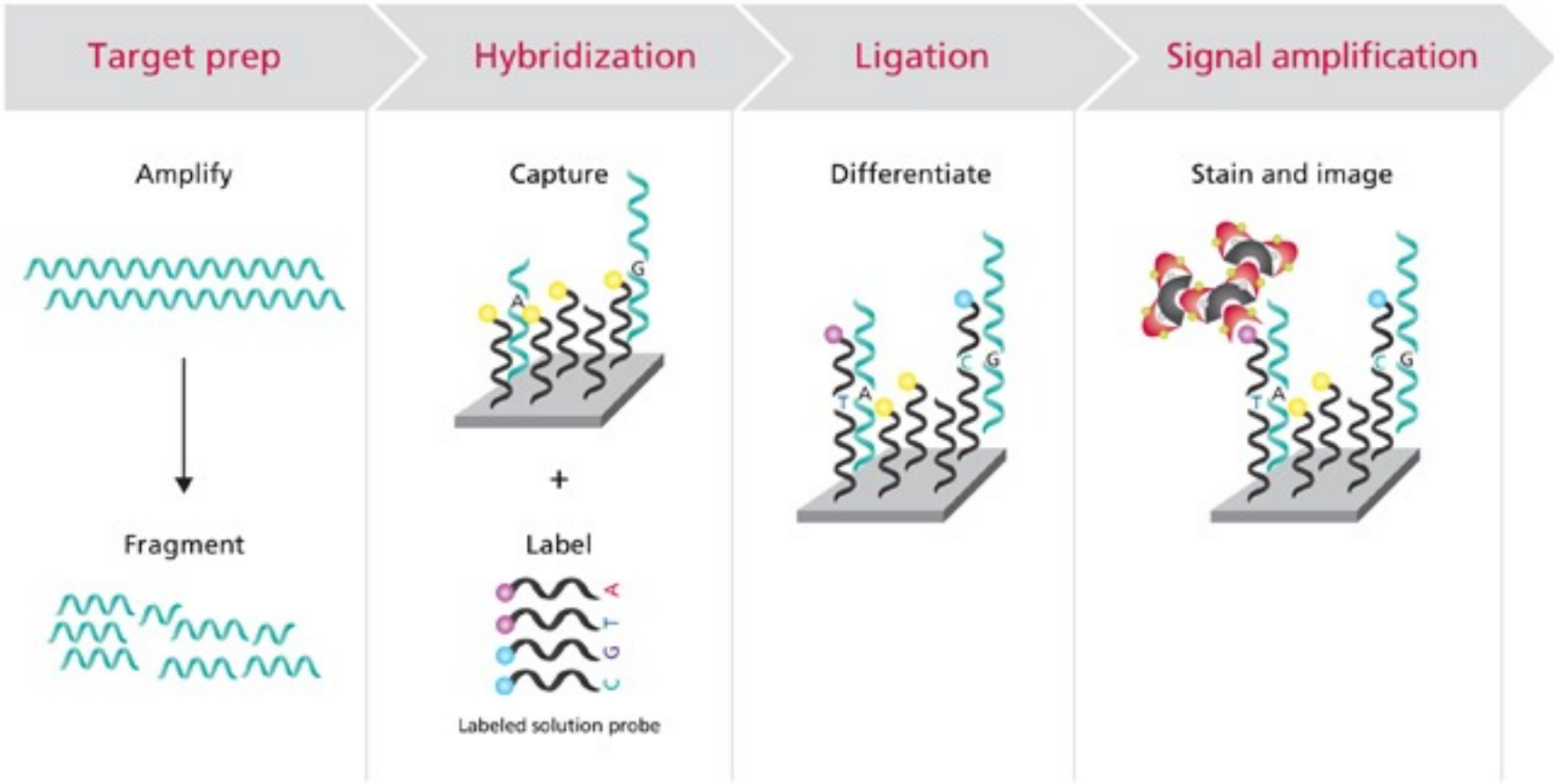
Affymetrix Axiom GW

Illumina Human Hap 650v3

Illumina Human Hap 550v3

Illumina Human Hap 300v3

SNP arrays


Target prep
Hybridization
Ligation
Signal amplification
Amplify
Fragment
Capture
+
Label
Labeled solution probe
Differentiate
Stain and image

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Genome-wide association studies

DTU

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

Cases vs controls

- Obtain DNA from a disease group (e.g. asthma) and a control group
- For each individual: Run DNA on a SNP array measuring the genotypes at almost 1M loci
- If the group consists of 1,000 individuals, one obtains 1,000 x 1,000,000 genotypes
- Quality Control
- Loop over all SNPs, and identify those that are significantly more common among cases than controls
- Those SNPs are associated with the disease (in this study)
- Not necessarily causal

DTU

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

GWAS tools

PLINK
http://pngu.mgh.harvard.edu/~purcell/plink/

SNPTEST
http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html

DTU

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

Quality control

- Allele, genotypes frequencies, HWE tests
- Missing genotype rates
- Inbreeding, IBS and IBD statistics for individuals and pairs of individuals
- non-Mendelian transmission in family data
- Sex checks based on X chromosome SNPs
- Tests of non-random genotyping failure
- Ethnicity

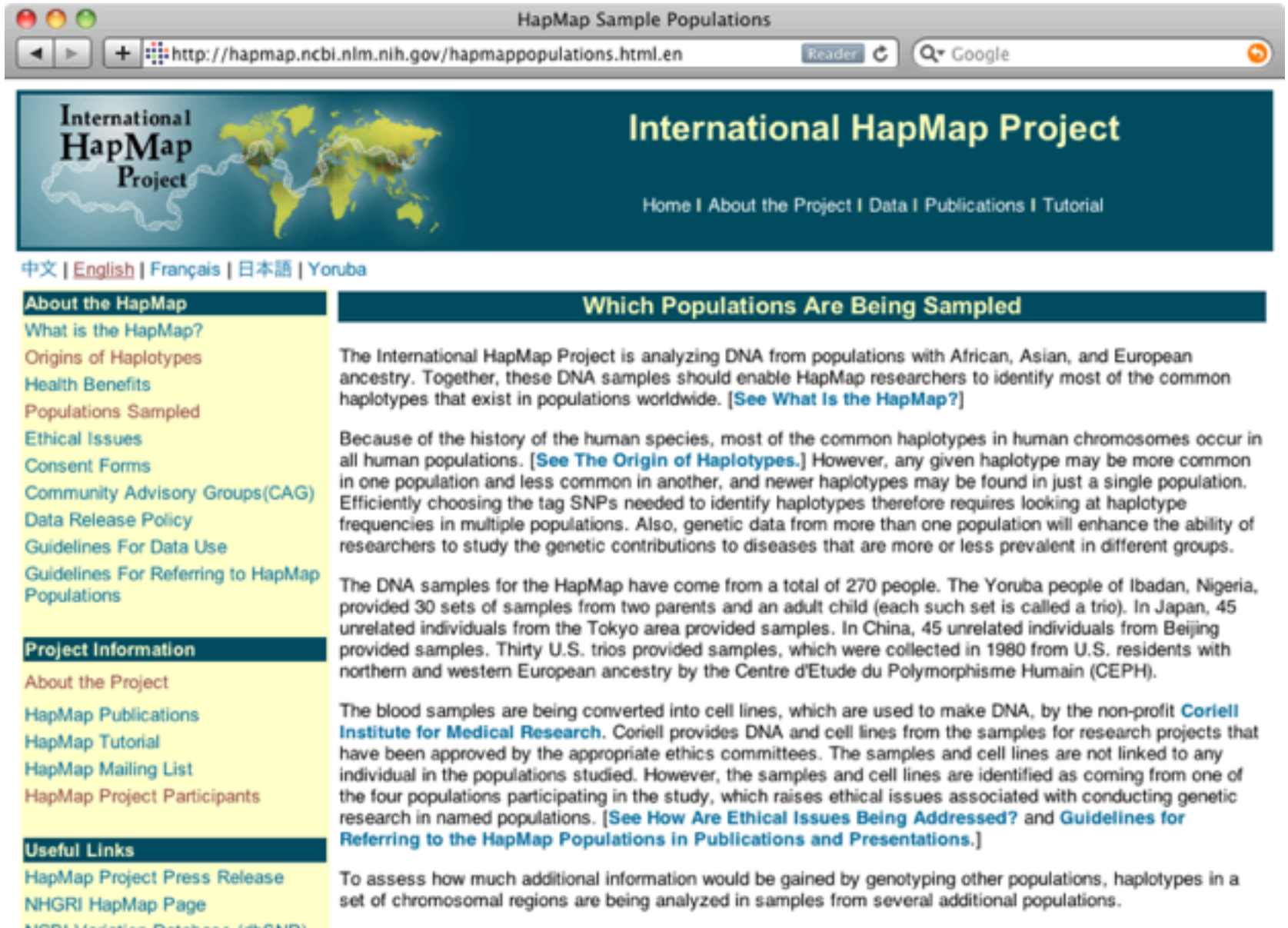
DTU

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

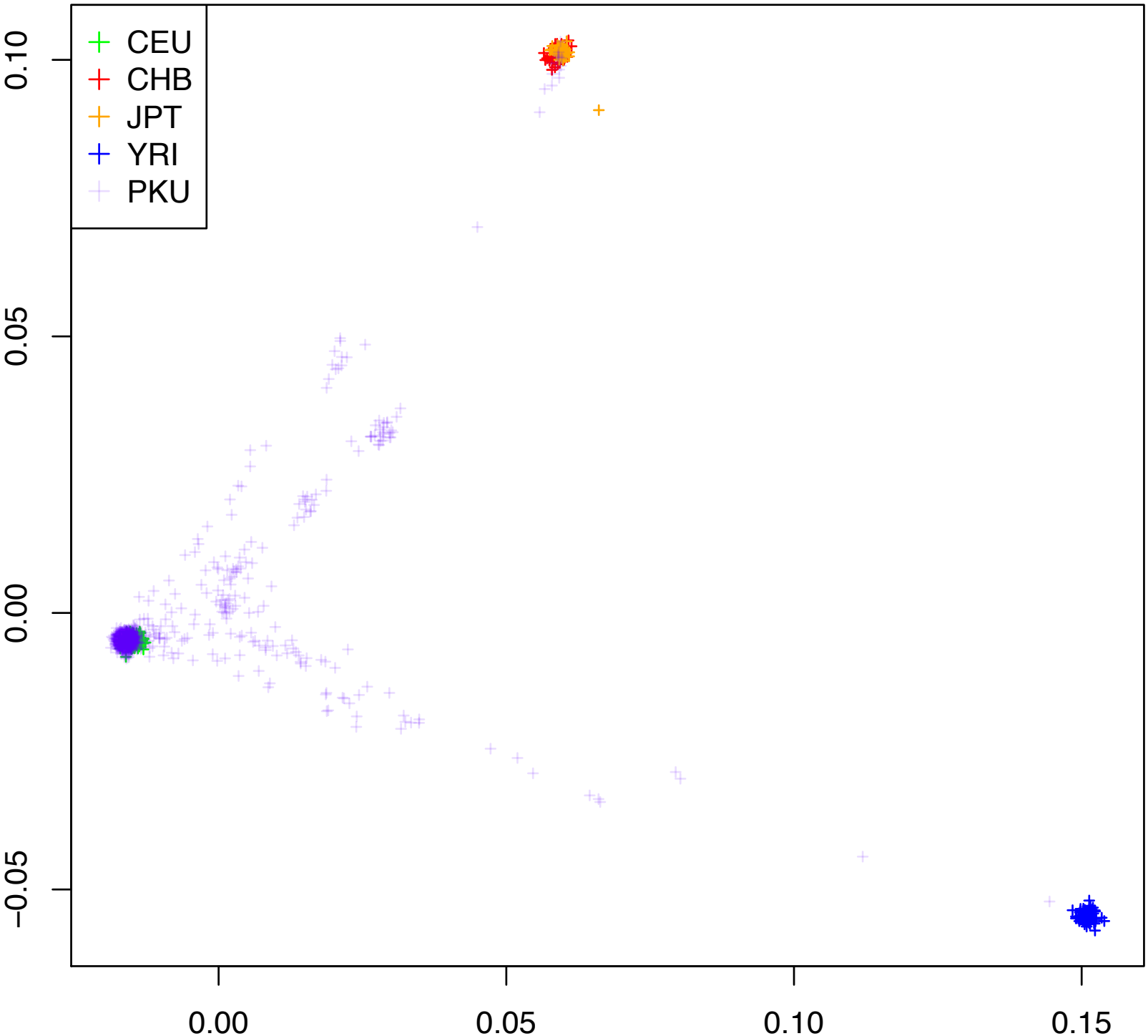
HapMap

- CEU
- CHB
- JPT
- YRI

...



PCA plot: patients vs HapMap



European Journal of Human Genetics (2010), 1–5

www.nature.com/ejhg

ARTICLE

NordicDB: a Nordic pool and portal for genome-wide control data

Monica Leu*[,1,2], Keith Humphreys[1], Ida Surakka[2,3], Emil Rehnberg[1], Juha Muilu[2], Päivi Rosenström[2], Peter Almgren[4], Juha Jääskeläinen[5], Richard P Lifton[6], Kirsten Ohm Kyvik[7], Jaakko Kaprio[2,8,9], Nancy L Pedersen[1], Aarno Palotie[2,10,11], Per Hall[1], Henrik Grönberg[1], Leif Groop[4], Leena Peltonen[2,3,10,11], Juni Palmgren[1,12] and Samuli Ripatti*[,2,3]

A cost-efficient way to increase power in a genetic association study is to pool controls from different sources. The genotyping effort can then be directed to large case series. The Nordic Control database, NordicDB, has been set up as a unique resource in the Nordic area and the data are available for authorized users through the web portal (http://www.nordicdb.org). The current version of NordicDB pools together high-density genome-wide SNP information from ~5000 controls originating from Finnish, Swedish and Danish studies and shows country-specific allele frequencies for SNP markers. The genetic homogeneity of the samples was investigated using multidimensional scaling (MDS) analysis and pairwise allele frequency differences between the studies. The plot of the first two MDS components showed excellent resemblance to the geographical placement of the samples, with a clear NW–SE gradient. We advise researchers to assess the impact of population structure when incorporating NordicDB controls in association studies. This harmonized Nordic database presents a unique genome-wide resource for future genetic association studies in the Nordic countries.

European Journal of Human Genetics advance online publication, 28 July 2010; doi:10.1038/ejhg.2010.112

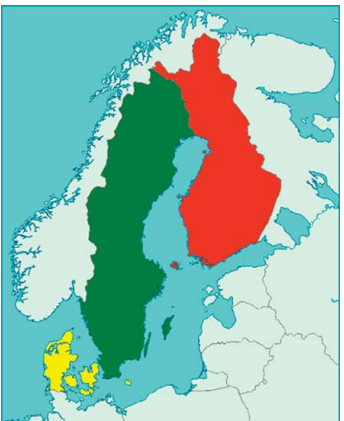


Figure 1 (**a**) Top axes of genetic variation in the Nordic Control Database, NordicDB (4620 samples) contrasted with the HapMap CEU (108 samples) and a Finnish HapMap reference population (81 samples). The MDS analysis was performed on ~45 000 SNPs that were common between genotyping platforms. The controls are part of the following studies: Cancer Prostate in Sweden (CAPS) 1 and 2, Cancer and Hormonal Replacement in Sweden (CAHRES), Diabetes Genetics Initiative in Western Finland and Southern Sweden (DGI-FIN and DGI-SWE), SGENE and MS in the Helsinki region, Aneurysm study in the Helsinki region, GenomEUtwin Denmark (GenomEUtwin-DK), GenomEUtwin Sweden (GenomEUtwin-SWE) and GenomEUtwin Finland (GenomEUtwin-FIN). (**b**) Geographical map of Scandinavia with three countries highlighted to show the origin of the samples in panel **a**: Finland (red), Sweden (green) and Denmark (yellow).

GWAS results: p-values

Example of GWAS results (asthma)

Manhattan plot displays all SNPs on x-axis (order by genomic location), and -$\log_{10}$ of p-values on y-axis.

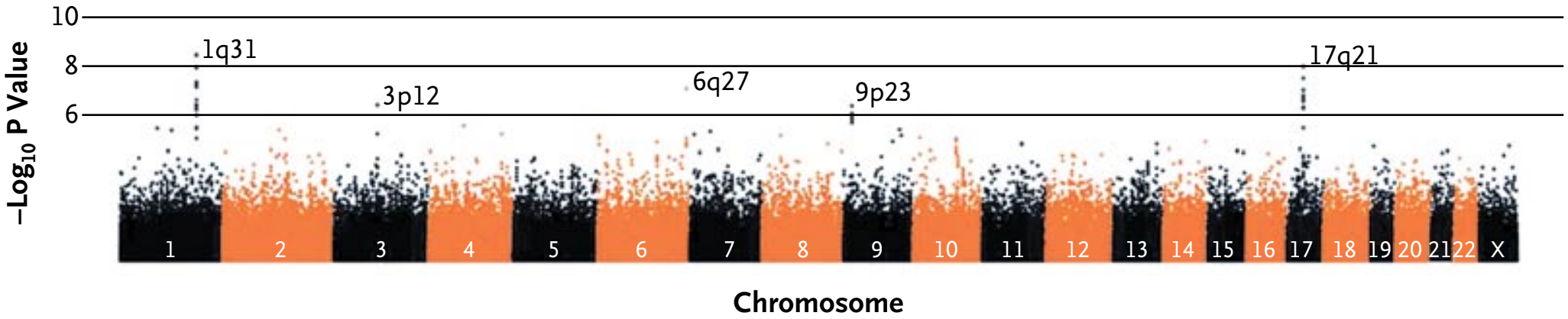


Figure 2. Manhattan Plot of the Results from the Combined Subjects of European Ancestry Who Had Asthma.

The $-\log_{10}$ P values are plotted against the physical distance. Only the two loci at chromosome 1q31 and 17q21 were significantly associated with asthma after Bonferroni correction. Individual chromosome labels are indicated in white within the Manhattan plot.

Variants of *DENND1B* Associated with Asthma in Children

Patrick M.A. Sleiman, Ph.D., James Flory, Ph.D., Marcin Imielinski, M.D., Ph.D., Jonathan P. Bradfield, B.S., Kiran Annaiah, M.Sc., Saffron A.G. Willis-Owen, Ph.D., Kai Wang, Ph.D., Nicholas M. Rafaels, M.S., Sven Michel, Ph.D., Klaus Bonnelykke, M.D., Ph.D., Haitao Zhang, Ph.D., Cecilia E. Kim, B.A., Edward C. Frackelton, B.A., Joseph T. Glessner, M.Sc., Cuiping Hou, M.Sc., F. George Otieno, M.Sc., Erin Santa, B.A., Kelly Thomas, B.A., Ryan M. Smith, B.A., Wendy R. Glaberson, B.A., Maria Garris, B.A., Rosetta M. Chiavacci, B.S.N., Terri H. Beaty, Ph.D., Ingo Ruczinski, Ph.D., Jordan M. Orange, M.D., Ph.D., Julian Allen, M.D., Jonathan M. Spergel, M.D., Ph.D., Robert Grundmeier, M.D., Ph.D., Rasika A. Mathias, Sc.D., Jason D. Christie, M.D., Erika von Mutius, M.D., William O.C. Cookson, M.D., Michael Kabesch, M.D., Miriam F. Moffatt, Ph.D., Michael M. Grunstein, M.D., Ph.D., Kathleen C. Barnes, Ph.D., Marcella Devoto, Ph.D., Mark Magnusson, M.D., Hongzhe Li, Ph.D., Struan F.A. Grant, Ph.D., Hans Bisgaard, M.D., and Hakon Hakonarson, M.D., Ph.D.

N Engl J Med 2010; 362:36-44 | January 7, 2010

Example of GWAS results (asthma)

Manhattan plot displays all SNPs on x-axis (order by genomic location), and -$\log_{10}$ of p-values on y-axis.

Figure 2. Manhattan Plot of the Results from the Combined Subjects of European Ancestry Who Had Asthma.

The $-\log_{10}$ P values are plotted against the physical distance. Only the two loci at chromosome 1q31 and 17q21 were significantly associated with asthma after Bonferroni correction. Individual chromosome labels are indicated in white within the Manhattan plot.

Variants of *DENND1B* Associated with Asthma in Children

Patrick M.A. Sleiman, Ph.D., James Flory, Ph.D., Marcin Imielinski, M.D., Ph.D., Jonathan P. Bradfield, B.S., Kiran Annaiah, M.Sc., Saffron A.G. Willis-Owen, Ph.D., Kai Wang, Ph.D., Nicholas M. Rafaels, M.S., Sven Michel, Ph.D., Klaus Bonnelykke, M.D., Ph.D., Haitao Zhang, Ph.D., Cecilia E. Kim, B.A., Edward C. Frackelton, B.A., Joseph T. Glessner, M.Sc., Cuiping Hou, M.Sc., F. George Otieno, M.Sc., Erin Santa, B.A., Kelly Thomas, B.A., Ryan M. Smith, B.A., Wendy R. Glaberson, B.A., Maria Garris, B.A., Rosetta M. Chiavacci, B.S.N., Terri H. Beaty, Ph.D., Ingo Ruczinski, Ph.D., Jordan M. Orange, M.D., Ph.D., Julian Allen, M.D., Jonathan M. Spergel, M.D., Ph.D., Robert Grundmeier, M.D., Ph.D., Rasika A. Mathias, Sc.D., Jason D. Christie, M.D., Erika von Mutius, M.D., William O.C. Cookson, M.D., Michael Kabesch, M.D., Miriam F. Moffatt, Ph.D., Michael M. Grunstein, M.D., Ph.D., Kathleen C. Barnes, Ph.D., Marcella Devoto, Ph.D., Mark Magnusson, M.D., Hongzhe Li, Ph.D., Struan F.A. Grant, Ph.D., Hans Bisgaard, M.D., and Hakon Hakonarson, M.D., Ph.D.

N Engl J Med 2010; 362:36-44 | January 7, 2010

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Online phenotype association resources

- NHGRI's Catalog of Published GWAS:

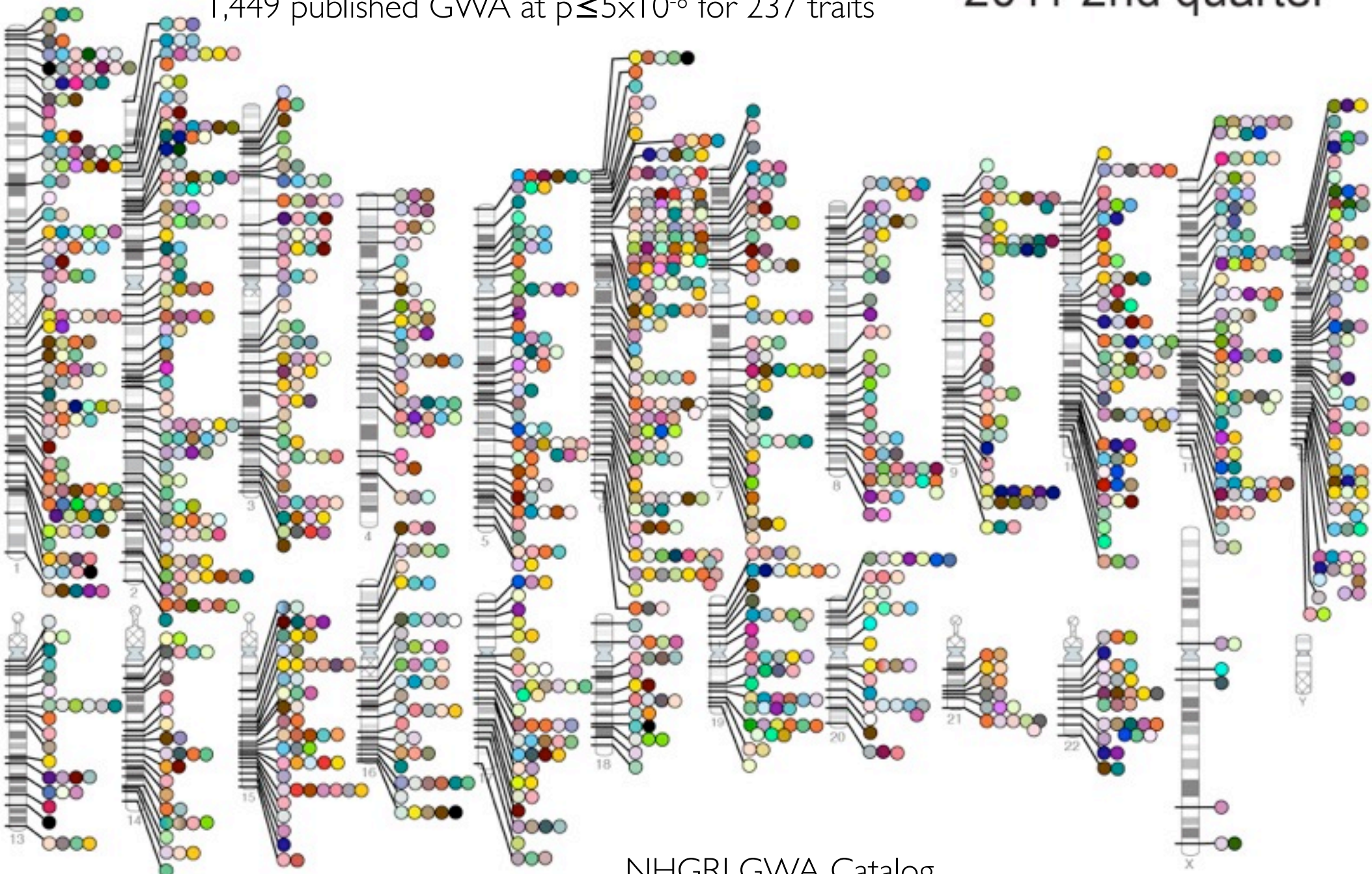
  http://www.genome.gov/GWAStudies/

- SNPedia:

  http://www.snpedia.com

DTU

Published Genome-Wide Associations through 06/2011,
1,449 published GWA at $p \leq 5 \times 10^{-8}$ for 237 traits



NHGRI GWA Catalog
www.genome.gov/GWAStudies


Abdominal aortic aneurysm
Acute lymphoblastic leukemia
Adhesion molecules
Adiponectin levels
Age-related macular degeneration
AIDS progression
Alcohol dependence
Alopecia areata
Alzheimer disease
Amyloid A levels
Amyotrophic lateral sclerosis
Angiotensin-converting enzyme activity
Ankylosing spondylitis
Arterial stiffness
Asparagus anosmia
Asthma
Atherosclerosis in HIV
Atrial fibrillation
Attention deficit hyperactivity disorder
Autism
Basal cell cancer
Behcet's disease
Bipolar disorder
Biliary atresia
Bilirubin
Bitter taste response
Birth weight
Bladder cancer
Bleomycin sensitivity
Blond or brown hair
Blood pressure
Blue or green eyes
BMI, waist circumference
Bone density
Breast cancer
C-reactive protein
Calcium levels
Cardiac structure/function
Cardiovascular risk factors
Carnitine levels
Carotenoid/tocopherol levels
Celiac disease
Celiac disease and rheumatoid arthritis
Cerebral atrophy measures
Chronic lymphocytic leukemia
Chronic myeloid leukemia
Cleft lip/palate
Coffee consumption
Cognitive function
Conduct disorder
Colorectal cancer
Corneal thickness
Coronary disease
Creutzfeldt-Jakob disease
Crohn's disease
Crohn's disease and celiac disease
Cutaneous nevi
Cystic fibrosis severity
Dermatitis
DHEA-s levels
Diabetic retinopathy
Dilated cardiomyopathy
Drug-induced liver injury
Drug-induced liver injury (amoxicillin-clavulanate)
Endometrial cancer
Endometriosis
Eosinophil count
Eosinophilic esophagitis
Erectile dysfunction and prostate cancer treatment
Erythrocyte parameters
Esophageal cancer
Essential tremor
Exfoliation glaucoma
Eye color traits
F cell distribution
Fibrinogen levels
Folate pathway vitamins
Follicular lymphoma
Fuch's corneal dystrophy
Freckles and burning
Gallstones
Gastric cancer
Glioma
Glycemic traits
Hair color
Hair morphology
Handedness in dyslexia
HDL cholesterol
Heart failure
Heart rate
Height
Hemostasis parameters
Hepatic steatosis
Hepatitis
Hepatocellular carcinoma
Hirschsprung's disease
HIV-1 control
Hodgkin's lymphoma
Homocysteine levels
Hypospadias
Idiopathic pulmonary fibrosis
IFN-related cytopeni
IgA levels
IgE levels
Inflammatory bowel disease
Insulin-like growth factors
Intracranial aneurysm
Iris color
Iron status markers
Ischemic stroke
Juvenile idiopathic arthritis
Keloid
Kidney stones
LDL cholesterol
Leprosy
Leptin receptor levels
Liver enzymes
Longevity
LP (a) levels
LpPLA(2) activity and mass
Lung cancer
Magnesium levels
Major mood disorders
Malaria
Male pattern baldness
Mammographic density
Matrix metalloproteinase levels
MCP-1
Melanoma
Menarche & menopause
Meningococcal disease
Metabolic syndrome
Migraine
Moyamoya disease
Multiple sclerosis
Myeloproliferative neoplasms
Myopia (pathological)
N-glycan levels
Narcolepsy
Nasopharyngeal cancer
Natriuretic peptide levels
Neuroblastoma
Nicotine dependence
Obesity
Open angle glaucoma
Open personality
Optic disc parameters
Osteoarthritis
Osteoporosis
Otosclerosis
Other metabolic traits
Ovarian cancer
Pancreatic cancer
Pain
Paget's disease
Panic disorder
Parkinson's disease
Periodontitis
Peripheral arterial disease
Personality dimensions
Phosphatidylcholine levels
Phosphorus levels
Photic sneeze
Phytosterol levels
Platelet count
Polycystic ovary syndrome
Primary biliary cirrhosis
Primary sclerosing cholangitis
PR interval
Progranulin levels
Progressive supranuclear palsy
Prostate cancer
Protein levels
PSA levels
Psoriasis
Psoriatic arthritis
Pulmonary funct. COPD
QRS interval
QT interval
Quantitative traits
Recombination rate
Red vs.non-red hair
Refractive error
Renal cell carcinoma
Renal function
Response to antidepressants
Response to antipsychotic therapy
Response to carbamazepine
Response to clopidogrel therapy
Response to hepatitis C treat
Response to interferon beta therapy
Response to metaformin
Response to statin therapy
Restless legs syndrome
Retinal vascular caliber
Rheumatoid arthritis
Ribavirin-induced anemia
Schizophrenia
Serum metabolites
Skin pigmentation
Smoking behavior
Speech perception
Sphingolipid levels
Statin-induced myopathy
Stroke
Sudden cardiac arrest
Suicide attempts
Systemic lupus erythematosus
Systemic sclerosis
T-tau levels
Tau AB1-42 levels
Telomere length
Testicular germ cell tumor
Thyroid cancer
Thyroid volume
Tooth development
Total cholesterol
Triglycerides
Tuberculosis
Type 1 diabetes
Type 2 diabetes
Ulcerative colitis
Urate
Urinary albumin excretion
Urinary metabolites
Uterine fibroids
Venous thromboembolism
Ventricular conduction
Vertical cup-disc ratio
Vitamin B12 levels
Vitamin D insufficiency
Vitiligo
Warfarin dose
Weight
White cell count
White matter hyperintensity
YKL-40 levels

Mini-exercise

Go this web-site:

NHGRI's Catalog of Published GWAS

http://www.genome.gov/GWAStudies/

1. Look around !
2. Search by
   - Disease/Trait: Type 2 diabetes
   - P-value treshold: 20 (=10^{-20})
3. Common top SNPs ?
4. Common genes (but different SNPs) ?

(Systematic approach: Meta-analysis, Imputation)

First Author/ Date/ Journal/ Study	Disease /Trait	Initial Sample Size	Replication Sample Size	Region	Reported Gene(s)	Strongest SNP-Risk Allele	Risk Allele Frequency in Controls	P-value	OR or beta-coefficient and [95% CI]	Platform [SNPs passing QC]
Voight June 27, 2010 *Nat Genet*	Type 2 diabetes	8,130 European descent cases, 38,987 European descent controls	Up to 34,412 European descent cases, 59,925 European descent controls	10q25.2 11q13.4 6p22.3	*TCF7L2* *CENTD2* *CDKAL1*	rs7903146-T rs1552224-A rs10440833-A		2×10^{-51} 1×10^{-22} 2×10^{-22}	1.4 [1.34-1.46] 1.14 [1.11-1.17] 1.25 [1.20-1.31]	Affymetrix & Illumina [2,426,886] (imputed)
Takeuchi April 29, 2009 *Diabetes*	Type 2 diabetes	519 Japanese cases, 503 Japanese controls	5,629 Japanese cases, 7,370 Japanese controls	9p21.3 11p15.5	*CDKN2A, CDKN2B* *KCNQ1*	rs2383208-A rs2237892-C	0.55 0.59	2×10^{-29} 1×10^{-26}	1.34 [1.27-1.41] 1.33 [1.27-1.41]	Illumina [482,625]
Timpson December 03, 2008 *Diabetes*	Type 2 diabetes	1,924 cases, 2,938 controls	3,757 cases, 5,346 controls	10q25.2	*TCF7L2*	rs7903146-?	NR	9×10^{-30} (non-obese)	1.49 [1.39-1.59]	Affymetrix [393,453]
Yasuda August 17, 2008 *Nat Genet*	Type 2 diabetes	187 Japanese cases, 1,504 Japanese controls	6,552 Asian cases, 6,621 Asian controls, 2,830 cases, 3,740 controls (Swedish)	11p15.5	*KCNQ1*	rs2237892-C	0.61	2×10^{-42}	1.4 [1.34-1.47]	Invader [82,343]
Zeggini March 30, 2008 *Nat Genet*	Type 2 diabetes	4,549 cases, 5,579 controls	24,194 cases, 55,598 controls	10q25.2	*TCF7L2*	rs7903146-T	NR	3×10^{-23}	1.37 [1.28-1.47]	Affymetrix and Illumina [2,202,892] (imputed)
Saxena April 26, 2007 *Science*	Type 2 diabetes	1,464 cases, 1,467 controls	5,065 cases, 5,785 controls (also includes meta-analysis from DGI +FUSION +WTCCC)	10q25.2	*TCF7L2*	rs7903146-T	0.26	2×10^{-31}	1.38 [1.31-1.46]	Affymetrix [386,731]
Zeggini April 26, 2007 *Science*	Type 2 diabetes	1,924 cases, 2,938 controls	3,757 cases, 5,346 controls (also includes meta-analysis from DGI +FUSION +WTCCC)	10q25.2	*TCF7L2*	rs7901695-C	NR	1×10^{-48} (DGI +FUSION +WTCCC)	1.37 [1.31-1.43]	Affymetrix [393,453]
Sladek February 11, 2007 *Nature*	Type 2 diabetes	661 cases, 614 controls	2,617 cases, 2,894 controls	10q25.2	*TCF7L2*	rs7903146-T	0.30	2×10^{-34}	1.65 [1.28, 2.02]	Illumina [392,935]

SNPedia

Earwax

rs17822931 determines wet vs dry earwax [PMID 16444273]

This can also be used to distinguish asian ancestry.

NCBI coffeebreak introduction

Category: Is a medical condition

Navigation

SNPedia
FAQ
Blog
Recent changes
Random page

Toolbox

What links here
Related changes
Special pages
Printable version
Permanent link
Browse properties

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Commercial genotyping companies

- 23andme
- deCODEme
- Navigenics

~ 1 million SNPs